INSTITUTION: Astrophysical Research Consortium
NSF PROGRAM: Special Projects-Division of Astronomical Sciences (AST)
PRINCIPAL INVESTIGATOR:
PROPOSAL TITLE: An Extension to the Sloan Digital Sky Survey

PANEL SUMMARY:
SDSS-II Reverse Site Visit

Panel Report (mod Feb 24; 11:55 AM MST)

Introduction

The Sloan Digital Sky Survey (SDSS) has not only made substantial contributions to a number of areas of astronomy, including cosmology, large scale structure, formation and evolution of galaxies, galactic structure, and stellar structure, it has also changed the way astronomy is done by a significant fraction of the community. It has provided one of the fundamental datasets around which development of the National Virtual Observatory has proceeded. In terms of high-impact publications, it has quickly reached a level that surpasses many space observatories.

It is not surprising that, as its initial operations period reaches its conclusion, the participants are thinking about how to best use the system that they have developed. This system is much more extensive and complex than a simple observatory; it includes an efficient imaging system, a sophisticated suite of data reduction pipelines, and a data archive and distribution system that supports a worldwide clientele. The system is more than hardware and software; it has now reached the stage of maturity at which the elements of its operation - acquire images, process images, select targets, drill plug plates, acquire spectra, process spectra, distribute data - take place smoothly and efficiently.

The proposal that they have written has three components: (1) complete the initial extragalactic survey, termed the Legacy component, for which the imaging will be complete, but the spectroscopy will have gap amounting to about 20% of the survey remaining, (2) undertake a new imaging/spectroscopic survey of the Galaxy's halo and thick disk, termed SEGUE, and (3) undertake a program aimed at discovering and measuring light curves for about 200 supernovae at redshifts between about 0.1 and 0.3

These three components are separable scientifically, but they are not separable programmatically. Because of the distribution of fields on the sky, the observations for the three projects fit together into a complete program that roughly subscribes the available observing time over the whole year. Thus, despite our discussion below of the relative priorities of the three components, based on their intellectual merit, the panel agrees with the project teams assessment that there is little savings from curtailing any

of them. That does not mean, however, that there are no decisions to be made. The overestimation of the frequency of good weather and the underestimation of technical difficulties have been responsible for the incompleteness of the Legacy project, and the three-year duration of the proposal is driven by the time required to complete it, and to also make considerable progress on SEGUE and the SN program. The panel recommends that NSF continue to regard the SDSS endeavor in its original view: it is an experiment rather than an open-ended operation. The current suite of three experiments are fully justifiable, but progress must be made on all of them, and the Legacy program especially must be completed within the three-year time frame.

In writing this report, we want to compliment both the proposers and the NSF staff for their participation in the process. The sequence of proposal, written review, written questions and answers, and then the reverse site visit was an effective means for the panel to get a good insight into the complex issues. The work that the proposers put into extensive, thoughtful answers to the initial set of questions was very much appreciated, and led to the panel feeling comfortable that all the ground had been covered by the end of the reverse site visit. We also want to acknowledge the value of the pilot projects that SDSS undertook to explore some of the issues associated with each of the new projects, SEGUE and SN. The results of these pilot studies provided specific answers to a number of the panel's questions about how these projects would be carried out.

This report presents comments first on the NSF review criteria, intellectual merit and broader impact. It then describes a number of requirements that the panel believes must be met to give confidence that the NSF funding will be effectively used. Following this, we list several suggestions that we believe will help the SDSS team deal with some of the concerns identified. Finally, we comment on the budget for the program.

Intellectual Merit

The panel views all three projects as having very strong scientific merit. However, we are able to rank them in terms of scientific merit. In summary, we believe that the greatest scientific return will come from completion of the initial SDSS, the Legacy component of the new proposal. Following that, we assign the next highest merit to the SEGUE project. Finally, we judge the Supernova survey as lowest in scientific return, but still of significant scientific value. Detailed justification of this ranking is given below.

Legacy, the completion of the original SDSS

The SDSS Legacy program is intended to complete the original objectives of the SDSS. SDSS I, when it terminates in June 2005 will fall short of these objectives in both a quantitative and qualitative sense. Although the imaging coverage is expected to be mostly complete, the spectroscopy will not. Originally intended to gather 1,000,000 galaxy spectra, the final SDSS I data release will have achieved about 80% of this. Crucially, however, the distribution on the sky of the missing 20% is concentrated in a

localized region rather than being uniform, random, or sparse. This has several significant negative implications for science based on SDSS.

The most obvious of these is on the power spectra derived from the galaxy distribution. The presence of the spatial gap broadens the windowing function, which increases the uncertainties in the derived power spectrum at short wavelengths. Additionally, the power spectrum lacks information at the longest wavelength scales. Both of these effects make comparison with cosmological models less fruitful. There is broad consensus in the cosmological community that removing this limitation is well worth doing, and we concur with this view.

The completion of SDSS will also have broader impact on astronomy. Increasingly, SDSS results are being used in conjunction with other surveys, to tie their photometry and astronomy to a well-defined system, and to make composite data sets that include data from optical and non-optical wavelengths. The SDSS is, in fact, becoming a crucial component of the NVO. In this context, the value of uniform coverage of a large, contiguous region of the sky is very great.

From the perspectives both of cosmology and of the general usefulness of the data through the NVO, our assessment is that completion of the SDSS is intellectually compelling.

SEGUE

SDSS-I has significantly advanced research in Galactic astronomy, especially through the discovery of stellar streams, and the identification of interesting classes of stars, from extremely metal-poor giants to T dwarfs.

The more methodical approach proposed in SDSS-II promises to enhance these discoveries and push the questions being raised to the next level. Specifically, the discovery and study of arcs in the Galactic sky aimed at identifying additional structures, or the lack thereof, has considerable promise. SDSS-II will remain a unique tool in the search for stellar streams and their implications for the formation of our Galaxy.

The large area to be scanned photometrically will lead to the identification of many interesting tracers for Galactic and stellar astrophysical research endeavors. The list of classes of stars for follow-up spectroscopy, including white dwarfs, G, K, and M dwarfs, blue horizontal branch stars, carbon stars, metal-poor dwarfs, and K giants, are all reasonable, but we assume the list will evolve. Clearly the coolest objects, potential T dwarfs, for example, or Mira variables, would be added to the list, as would the metal-rich equivalents to blue horizontal branch stars (red horizontal branch or red clump stars) as powerful tracers of the thick disk and thin disk populations.

The radial velocity precision goal of 7-10 km/sec appears feasible and is suitably

smaller than the one-dimensional velocity dispersion of the coolest stellar population, and is also very useful in distinguishing membership in or even the reality of stellar streams, be they remnants of dissolved clusters or accreted stellar ensembles. The USNO-B proper motions will supply valuable additional discrimination out to distances of 2-3 kpc for the disk populations.

The panel's concerns about handling the effects of reddening in target selection were mostly mollified through the careful selection of the planned observations, in tandem with consideration of the reddening maps of Schlegel et al. (1998). The panel is intrigued by the planned efforts to exploit the spectra of normal stars to assign temperatures, gravities, and metallicities, and encourages this work to proceed. In combination with the photometry, these results may provide a very valuable addition to our understanding of the origins of stellar streams and the Galaxy's major stellar populations.

The Supernova Survey

The proposed SN Ia survey seeks to shed more light on the nature of the Dark Energy responsible for the observed acceleration of the universe. It competes with two higher-redshift SN Ia surveys, SNLS (Supernovae Legacy Survey) and ESSENCE (Equation of State: SupErNovae trace Cosmic Expansion), which have already been taking data for 2+ years and claim to be able to measure w to a similar precision. For this reason, the panel does not consider the SN Ia survey to be a strong justification in itself for funding of SDSS-II. However, the SN Ia survey does mesh extremely well with the observing requirements of the Legacy and SEGUE surveys, and complements the SNLS and ESSENCE projects in at least two important ways. These are:

1. The SDSS-II SN Ia survey is unique in spanning a redshift range which overlaps both with ongoing low-redshift surveys like SNF (Supernova Factory) and CSP (Carnegie Supernova Program), and with the higher-redshift SNLS and ESSENCE projects. Hence, the problem of a systematic error creeping in due to differing photometric systems between the low- and high-redshift samples is eliminated. The fact that the SDSS photometric system is already uniformly calibrated and well-characterized is an additional plus.

2. By obtaining u- and g-band light curves for a significant number of SN Ia at redshifts $0.05 < z < 0.3$, it should be possible to study in more detail the effects of metallicity and age on the uv spectral energy distributions of SN Ia. These data will complement the u-band light curves that will be obtained by the CSP and SNF of $z < 0.10$ SN Ia, and the ultraviolet spectroscopic observations of low-z SN Ia currently being carried out with HST by Filippenko & collaborators.

The panel was impressed by the success of the pilot project carried out during the fall of 2004, and believes that the SDSS collaboration should be able to obtain well-sampled

ugriz light curves of at least ~150 SN Ia during the proposed 3 year extension. It also was pleased to see to see that new collaborations have been formed since the submission of the original SDSS-II proposal which will provide the ability to carry out spectroscopic follow-up at a level that will allow peculiar SN Ia to be recognized and K-corrections to be improved.

Broader Impacts

In the proposal and the reverse site visit, some impressive statistics are presented for outside scientific use of the SDSS-I data. The large number of papers written by scientists not associated with the SDSS team, the rapidly growing citation rate of SDSS-related papers, and the huge number of hits on the SDSS site all are metrics of the impact that SDSS is having on the scientific community and the public at large. SDSS has provided a substantial fraction of the ground-based optical/IR content around which the fledgling NVO has begun its development. The potential for bringing astronomical data and the tools with which to study it to researchers everywhere is overwhelming.

What will be the impact of SDSS-II? Regarding the Legacy program, it is fair to say that a separation of SDSS-I and SDSS-II would be artificial and difficult to accomplish. The value-added impact of the Legacy program to the original goals of the project is discussed above in the intellectual merit section. The scientific impact of what has become the Legacy program can be expected to continue to grow.

The SEGUE and SN programs should also have substantial scientific impact. The immediate goals of both projects are well-described in the proposal, and the proposal also calls out the potential follow-on studies that can be done. The broader impact of the SEGUE and SN programs rests in the future studies; Because of its vast size, the SEGUE data set is a potential gold/data mine.

Beyond the professional astronomical community, the impact is less-clearly stated in the proposal. A number of items are mentioned, and others were brought up in the reverse site visit. Several specific points were made: (1) The number of visits to the SDSS data site from the general public now exceeds the number from the astronomical community. (2) A series of student lab exercises have been written to involve SDSS. (3) Although these were not called out in detail, there are apparently a considerable number of individual EPO projects underway at the individual project institutions. (4) Finally, the collaboration with the American Museum of Natural History is an on-going and developing project.

Nevertheless, there is an apparent lack of a cohesive EPO strategy, and even an apparent lack of awareness of what is being done within the project. Some of these efforts involve the use of SDSS for outreach activities at the individual institutions; a few are collaboration-wide efforts. We believe more could be done by building EPO projects

around the notable strengths of the SDSS data and collaborators. In particular, we note the project's unique combination of understandable data, dramatic images, and a built-in means to provide universal access. The project would be well-served to develop such a strategy and to publicize what is already being done.

Overall Recommendation

The panel recommends that the NSF fund this proposal after mutual agreement is reached on the following issues:

1. The panel strongly recommends that, barring any unforeseen catastrophic failures during the proposed 3-year extension, the NSF require that a minimum acceptable outcome be established for SDSS-II, and that priorities be set so that this outcome is achieved. In particular, it is not acceptable in our view to say that the project will be defined as completed when the 3 years have expired. Rather, we recommend that completion be defined and measured in terms of an achievable metric.

It is the unanimous opinion of the panel members that in defining this metric, the completion of the Legacy survey should have the highest priority. Accepting that weather is a variable factor that cannot be controlled, it is vital that a strategy be developed for filling in the missing spectroscopic observations in a way that will allow the major scientific objectives of the survey to be achieved. In particular, the panel recommends that simulations of the type illustrated in Fig. 8 of the proposal be utilized to develop an observing strategy which a) achieves essentially all of the factor of three improvement in the minimum width of the window function of the power spectrum, even in the worst weather scenario, and b) permits significant progress in the SEGUE and SN I projects to be made to allow these programs to achieve the majority of their stated goals.

2. Both the SEGUE and SN programs will add new dimensions to SDSS data. In the case of SEGUE, stellar objects will have stellar parameters such as Teff, log g, and [Fe/H] as well as radial velocities. The SN program adds subtracted images, and most importantly, time dependence to SDSS data.

We believe it is crucial that both the SDSS data model, and the publicly available tools for accessing SDSS data, be expanded to incorporate the new forms of data from SEGUE and SN. The alternative, to leave some data as isolated data sets with inadequate access tools, will greatly reduce the broad impact of these programs. Accomplishing this will require significant work, particularly to incorporate time dependent data. We note that time dependent SDSS data that has already been taken is not publicly available, and there is apparently no plan in place to bring it into the SDSS data framework. The SDSS II should take note of the emerging VO standards for time dependent data and transient alerts, and incorporate them into a broadly useful data access portal.

3. The panel recommends that the project develop a cohesive EPO program to build on the present EPO efforts by the member institutions. This program might include the following elements:

o Developing an organizational awareness of the importance of EPO activities as a way to satisfy the broader impact requirement for NSF funding. This awareness will also help bring visibility of this important project to the public at large.

o Building on the relationship with the AMNH through additional programs, particularly those that might expand the program to a "road show" extending the educational benefits beyond New York. While there was some discussion of such activities described in the proposal and in the presentation to the panel, they were thin on details.

o Designate a person as EPO officer for the project. It is common for projects of this scale to have a dedicated EPO manager who can coordinate activities among the various institutions of the consortium, work with the AMNH and possibly other organizations, and bring increased visibility to the work already being done

4. The panel believes that the community's interests will best be served if a mechanism for annual reviews convened by the NSF is put in place for the duration of the extended SDSS mission. These reviews should consider progress and status in both scientific and management areas and should specifically include the modifications to the proposal discussed above.

Additional Recommendations

In addition to the requirements described above, the panel makes the following recommendations:

1. We recommend that the SDSS project explore whether the costs of operation on the mountain could be reduced by the large increase that has recently occurred in network bandwidth. Possibilities to consider would include the relocation of data quality assurance operations off the mountain, and transport of data via the network instead of by expensive media.

2. The panel was impressed with the significant enlargement in the community directly engaged in the planning for SDSS-II. If the funding contributed by these partners, plus the Sloan Foundation, and NSF significantly exceed the proposed budget, the panel agrees that including such funds to augment the proposed 2% contingency budget makes sense. Given the current smooth functioning of the Observatory, the data reductions, and the data archival, the original plan for a contingency, or management reserve, of 5%, appears sufficient. If the contingency exceeds that significantly, the panel recommends that management should prepare a plan for the disposition of remaining funds as SDSS-II nears its end. Suggestions that emerged during our discussions included funding to help the transition to curation of the SDSS-I and

SDSS-II data, and developing programs for further use of the telescope. Other uses may well occur to the partners.

3. The panel suggests that the SDSS-II team explore the development of tools to aid in telescope scheduling. Since the scheduling problem now involves the interplay among three projects, each of which will have their own independent strategy to maximize success, it is probably time to move past Excel spread sheets and pure thought. Tools that allow the exploration of different trades, both of the sequence of observations within programs and between programs, will allow all three teams to converge on an acceptable approach. Despite the success of the test last fall, it seems likely that decisions that were easy when the projects were just starting will be much more difficult near the end.

4. Although the tools that SDSS has developed for data access and mining have become much more powerful and robust, it is still the case that many astronomers are unaware of their capabilities or are put off by their apparent complexity. Also, there are a large number of subtle aspects of the data and its processing that researchers must know about in order to use the data effectively. Brian Yanny's brief demonstration of SDSS data access and manipulation was an excellent example of one of a set of presentations that could be made at an AAS meeting as a workshop on the use of SDSS data.

5. The panel acknowledges the impressive number of papers that use SDSS data written by both external collaborators and genuine outsiders. However, the panel suggests that the science teams for the three projects solicit expressions of interest from the broad community in becoming members of these teams. We would not force the collaboration to accept everyone who expresses interest, but we believe that the teams could be made stronger by casting the net more widely.

Budget/cost considerations

The panel remains a little uncomfortable with our understanding of the budget. The size and complexity of the program (including support for development, facility operations, and support for data processing and distribution) and the relatively high level of the data presented made it difficult to assess the ?comfort? level in detail. We tried to identify specific areas where we had concerns - contingency, observers - but, in the end, we rely on our overall impression that the budget is reasonable, primarily with the following evidence:

1. The annual cost for the period of performance is roughly equal to the annual cost for recent years of SDSS operations.

2. After a rocky start to SDSS in terms of schedule and budget, it appears that the collaboration has converged to an efficient management structure that works well. Recent data releases have been on time and consistent with promised capability.

3. The collaboration recently held an external review of their operations approach and budget, and this review resulted in a favorable outcome.

PANEL RECOMMENDATION:
PANEL RECOMMENDATION KEY:
DNF:Do Not Fund, FIP:Fund If Possible, F:Fund